CSCI6421: Distributed Systems Practice Problems

1. True or False

_____ Two programs communicating over sockets must be written in the same programing language.

_____ Two programs communicating over sockets must use the same transport protocol (TCP or UDP).

_____ Peer to Peer services are the only type of distributed system where a node is both a client and a server.

_____ Lamport clocks cannot always put all events into a total order, but vector clocks always can.

_____ Amazon's Elastic Map Reduce is a good example of an Infrastructure as a Service Cloud.

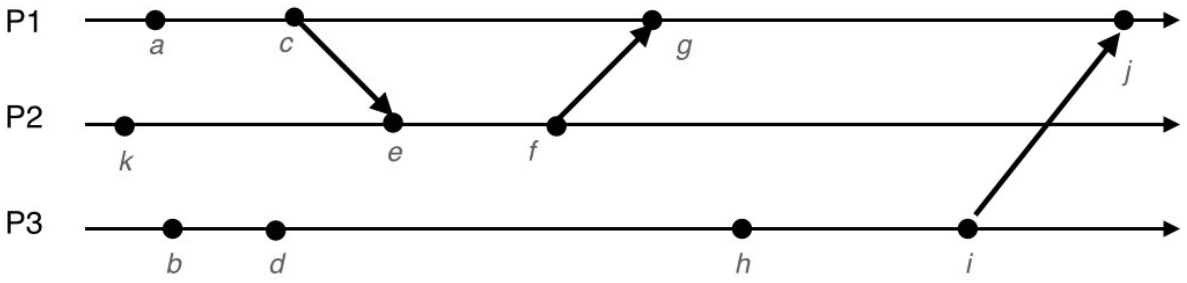
_____ Virtualization provides strong isolation, but it is generally slower than running code natively.

_____ Containers are a good choice for operating system kernel developers because they can be restarted very quickly, allowing you to debug as you modify the container's OS kernel.

_____ Most cloud applications use UDP to communicate with clients.

2. Clocks and Ordering

Annotate the process timeline with Vector Clocks for each event.



Using your vector clocks, indicate whether an event happens before (->), after (<-) or is concurrent (ll).

b _____ g g _____ i

j _____ h f _____ j

3. Fault Tolerance

You are designing a distributed system that uses replication to mask faults. If we assume at most **f=4** faults can ever happen at one time, then how many replicas are needed in each of these cases:

a) Replicas either respond correctly, do not respond at all, or return an arbitrary incorrect value.

b) Replicas either respond correctly or do not respond at all.

d) Replicas either respond correctly or return an error code: NULL.

4. Distributed Systems Challenges

Pick 3 of the distributed systems challenges below and write 3-4 sentences for each one explaining a system or technique we learned about that helps resolve them.

Challenges: heterogeneity, openness, security, fault handling, concurrency, quality of service, scalability, transparency

5. Cloud Types

Compare the scalability, ease of use, and security characteristics of an IaaS, a PaaS, and a private data center.

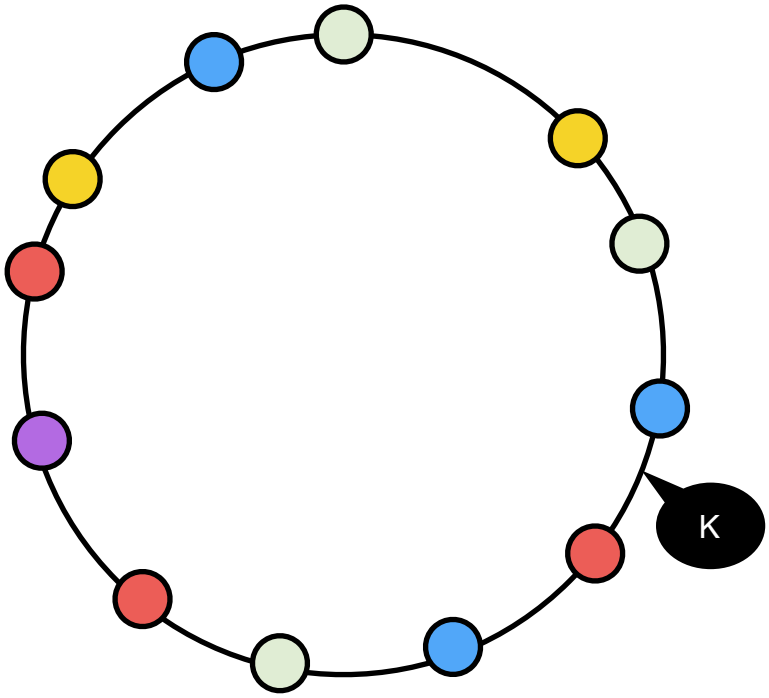
6. Virtualization

a) Why do virtual machines use technologies like shadow page tables? Why can't VMs be given real physical memory addresses to work with directly?

b) Why are containers "lighter weight" than VMs?

7. Distributed Hash Tables

Consider the DHT shown at right. Assume each node is responsible for keys from its starting point and proceeding clockwise around the ring to the next node.



If the purple node wants to get the data related to Key K with the indicated hash value, how many hops will it take to reach K if... (show the path for each case)

a) Each node is only connected to its immediate neighbors?

b) Each node is connected to its neighbors 1, 2, 4, and 8 hops away, moving clockwise around the ring?

c) What are the benefits and drawbacks of having every node connected to every other node in the ring?

d) The DHT we discussed assumed that nodes would announce when they were leaving, allowing them to copy their data to an adjacent node. Briefly describe would you have to modify this system if you wanted to support nodes crashing (i.e. leaving without announcing their departure).

8. SDN and NFV

a) Describe the steps that occur when an unknown flow arrives at an SDN-enabled switch.

b) Describe two of the techniques used by NFV systems to improve their performance and explain why they help.

9. Scaling

Suppose you have a stateful service. What would techniques could you use to scale it vertically or horizontally? (describe each individually)

10. Case Study

For this question you will be judged on the completeness, creativity, and clarity of your answer. Note that the case study description is intended to be somewhat vague. You should specify what assumptions you are making and how that affects your solution.

You are building a sensor network to study a mountain chain with several active volcanos. The network will be composed of two types of nodes: 1) Small nodes that have a radio with a radius of about 100 meters and are low power with limited processing and memory capabilities; 2) Large nodes that have a radio range radius of 1000 meters and have greater storage and processing capabilities than small nodes. The nodes are designed to be dropped out of an airplane over a wide area.

Each of the nodes will gather data that will be analyzed by scientists to detect potential eruptions. Both types are capable of gathering the same data from their sensors. A data point takes 1KB of memory. The small nodes have 10KB of total memory available and the large nodes have 1,000KB.

The scientists want to be able to answer questions like: what is the average data reading in a geographic region covered by several sensors? However, in the future they may devise more complicated programs that will run on each sensor, so they want to have a flexible system.

Draw diagrams and write an explanation to justify your design for this system. Which of the distributed systems challenges below does your system help with? How?
Challenges: heterogeneity, openness, security, fault handling, concurrency, quality of service, scalability, transparency